7-20-2005

When Should One Substract Background Fluorescence in Two Color Microarrays?

Robert B. Scharpf
Johns Hopkins Bloomberg School of Public Health, Department of Biostatistics, rscharpf@jhsph.edu

Christine A. Iacobuzio-Donahue
Johns Hopkins University School of Medicine, Department of Pathology, ciacobu@jhmi.edu

Julie B. Sneddon
Department of Biochemistry, Stanford University School of Medicine, sneddon@cmgm.stanford.edu

Giovanni Parmigiani
Johns Hopkins University, Department of Oncology and Department of Biostatistics, gp@jimmy.harvard.edu

When should one subtract background fluorescence in two color microarrays?

Robert B. Scharpf
Department of Biostatistics,
Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA.

Christine A. Iacobuzio-Donahue
Department of Pathology,
Johns Hopkins University School of Medicine, Baltimore, MD, USA.

Julie B. Sneddon
Department of Biochemistry,
Stanford University School of Medicine, Stanford, CA, USA.

Giovanni Parmigiani
Departments of Biostatistics[†] and Oncology[‡],
Johns Hopkins Bloomberg School of Public Health[†] and
Johns Hopkins University School of Medicine[‡], Baltimore, MD, USA.

Abstract

Two color microarrays are a powerful tool for genomic analysis, but have noise components that make inferences regarding gene expression inefficient and potentially misleading. Background fluorescence, whether attributable to non-specific binding or other sources, is an important component of noise. The decision to subtract fluorescence surrounding spots of hybridization from spot fluorescence has been controversial, with no clear criteria for determining circumstances that may favor, or disfavor, background subtraction. While it is generally accepted that subtracting background reduces bias but increases variance in the estimates of the

ratios of interest, no formal analysis of the bias-variance trade off of background subtraction has been undertaken. In this paper, we use simulation to systematically examine the bias-variance trade off under a variety of possible experimental conditions. Our simulation is based on data obtained from two self versus self microarray experiments and is free of distributional assumptions. Our results identify factors that are important for determining whether to background subtract, including the correlation of foreground to background intensity ratios. Using these results we develop recommendations for diagnostic visualizations that can help decisions about background subtraction.

1 Introduction

Two color microarrays evaluate the expression of thousands of genes and expressed sequence tags (EST's) in a single assay by quantifying the relative abundance of messenger RNA (mRNA). The discovery of differentially expressed genes using microarrays depends crucially on the choice of normalization [24, 26, 19]. Considerations for optimal normalization of the microarray data are platform-dependent.

The focus of this paper is on noise intrinsic to fluorescent imaging platforms. Specifically, we consider cDNA microarrays where target and reference mRNA are reverse-transcribed to cDNA and tagged by green and red fluorophores. The target and reference preparations are combined and competitively hybridized to short DNA sequences (probes) spotted on a glass slide. Each probe on the array binds, in theory, to a single gene or EST. After imaging the array, statistics such as the median red and green intensity at each spot (foreground) as well as comparable statistics for the local fluorescence surrounding the spot of hybridization are usually available. We will refer to the latter measure of fluorescence as background. Estimates of background can be highly variable and are sensitive to the imaging methodology used [3]. Background is often subtracted from foreground prior to normalization. Ideally, the added variability would be compensated by a reduction in bias. See [22, 12, 25] for a more complete description of cDNA microarray technology.

Background can arise from a number of sources, including incomplete washing after hybridization, features of the slide that bind dye or RNA, and imprecision in spot localization (segmentation) during image acquisition. See [23] for a comprehensive list of sources of variability in cDNA microarrays. Background subtraction (BS) is an imperfect solution for reducing bias due, in part, to imprecision of the imaging measure of background, as well as heterogeneity of background near the spot of hybridization [3]. BS introduces

another layer of variability to the gene expression measure.

The decision to implement BS plays an important role in identifying differentially expressed genes. See [4, 1, 21] for considerations when inferring differential expression by ratios of signal intensity. Subtracting background from low abundance genes results in overdispersion of $\log$ intensity ratios. Also problematic with low abundance genes is the potential for estimates of background to be greater than foreground. We and others believe that subtracting local estimates of background from foreground is less than ideal [3, 20, 16]. More sophisticated normalization methods have been implemented to deal with this problem [15, 3]. Nevertheless, the decision of whether to perform BS has been largely a matter of personal preference with few guidelines for determining when BS is appropriate. One barrier to a more formal analysis of the bias-variance trade off has been the absence of a suitable model for simulating the variability in microarray experiments.

Factors influencing the bias and variability in microarray data are not limited to the abundance of cDNA in the hybridized samples. Implicitly, BS assumes that the background is homogeneous across spotted and nonspotted portions of the array. However, this assumption is often not valid. Foreground fluorescence arising from cross-hybridization (whether specific or nonspecific) and location-specific binding are common and each contribute to unmeasured background heterogeneity. BS is inappropriate if such location-specific biases exist [16, 3]. Location-dependent normalization procedures such as loess (see Section 2) may only partially correct for this problem. Diagnostics for visualizing when such biases are likely to exist are needed.

Because our interest lies in the ratio of red and green intensities, we advocate the correlation of the foreground ratio to the background ratio as a diagnostic. Specifically, a high correlation of the ratios of foreground to background suggests a bias of the foreground

ratio that is not driven by complementarity of nucleotide sequences. Conversely, a low correlation of the foreground to background ratios suggests that the foreground ratios vary due to differences in hybridized transcript. If one is to use the correlation of ratios as a diagnostic, the question then becomes at what level of correlation will the benefit in bias reduction by BS compensate for the introduced variability. We verify the importance of correlation on the bias-variance trade off through simulation in Section 3 and discuss additional diagnostics for measuring the correlation of the ratios in subregions of the microarray in Section 4.

In addition to the correlation of foreground to background ratios, our simulation considers multiple factors that are likely to influence the decision to perform BS, including the abundance of hybridized transcript. We use two self versus self (SVS) microarray experiments (described in Section 2) for the simulation to insure that our results are not biased to the technological variability in one experiment. Advantages of using SVS experimental data include that the true differential expression is known to be absent, variability in the gene expression is from actual data, and thousands of genes can be simulated, rather than one at a time. Because we compare the bias, variance, and mean squared error (MSE) with and without BS, these results provide guidance on whether to subtract estimates of background from foreground in two color microarrays.

2 Methods

R [14] and Bioconductor (http://www.bioconductor.org) [10] were used for all statistical analyses. Our simulation uses data from two SVS hybridizations. SVS hybridization was performed by labeling one aliquot of cDNA with red fluorophores and a separate aliquot from the same sample with green fluorophores. Equal mass amounts of red- and green-labeled cDNA were then competitively hybridized to the microarray. In truth, no differential expression should occur.

Dataset 1 A SVS hybridization of amplified Stratagene universal reference RNA was obtained from the Stanford Microarray Database (http://genome-www.stanford.edu/microarray) [11]. Stratagene human universal reference RNA contains RNA from ten pooled human cell lines. Universal reference RNA is commonly used as the reference sample in microarray platforms that use competitive hybridization to quantify relative mRNA abundance since most arrayed genes in the pooled sample are detectable above background noise [27]. The microarray for this experiment was spotted with 43,104 clones. Background fluorescence was computed as the median pixel intensity from several locations adjacent to the spot of hybridization. Similarly, spot fluorescence was computed as the median pixel intensity from several locations within the target region for hybridization. Segmentation and image analysis was performed using GenePix Pro 3.0.6.86 (Axon Instruments, Inc.). See [9] for more information on GenePix. This data is publicly available at the Stanford Microarray Database.

Dataset 2 A second SVS hybridization of breast cancer cell line MCF7 was downloaded from supplementary material at http://www.ece.ucsb.edu/pubs/bmc [2]. The microarray

was scanned with an Agilent laser confocal scanner and gridded using the DEARRAY software [5]. Spot and background fluorescence were calculated as average intensities within the target area, after trimming the top and bottom 5%. The cDNA microarray was printed with 11,520 clones from Incyte Genomics and 1136 clones from Research Genetics library for a total of 13,440 spots. See [2] for more detailed information regarding this experiment.

We hereafter refer to datasets 1 and 2 as *Stratagene* and *MCF7*, respectively.

2.1 Filtering

Negative spot intensities after background subtraction are not sensible and methods that use background subtraction typically exclude such spots. To facilitate comparison of BS to no background subtraction (NBS), we excluded spots where background was measured greater than foreground (though this is typically not necessary for NBS). 28,837 of 43,104 spots and 6933 of 13,440 spots had foreground greater than background in both channels for the Stratagene and MCF7 experiments, respectively. Because pixel level data within a spot has been shown to be a useful indicator of spot quality [3], additional filtering criteria were applied to the Stratagene dataset to obtain a smaller subset of 16,908 spots. The additional filtering required a correlation greater than 0.6 of red and green pixels within a spot, no flags generated from the GenePix imaging software, and median foreground 1.5 fold greater than median background for both the red and green intensities. The simulation was performed with minimal filtering of the data (28,837 and 6933 spots for Stratagene and MCF7, respectively), as well as with the more filtered Stratagene subset (16,908 spots).

2.2 Normalization

The spot statistics used for normalizing the microarrays are the $\log_2$ abundance (A) and the $\log_2$ ratio (M) of median red (R) and green (G) foreground. Hence, $A, M, R,$ and G are spot statistics computed without subtracting background. M_s and A_s are the corresponding statistics for the ratio and abundance, respectively, after subtracting median red (R_b) and green (G_b) background. Explicitly,

$$\begin{aligned} A &= \frac{1}{2}\log_2(RG) \\ A_s &= \frac{1}{2}\log_2\left[(R-R_b)(G-G_b)\right] \\ M &= \log_2\left(\frac{R}{G}\right) \\ M_s &= \log_2\left(\frac{R-R_b}{G-G_b}\right). \end{aligned}$$

A-dependent normalization was performed by robust locally weighted least squares regression (loess) [6, 7] using Bioconductor software [10, 8]. A-dependent normalization procedures for smoothing MA scatterplots are often preferable to global-normalization due to the frequent occurrence of intensity biases [18]. In addition, we used loess to smooth scatterplots of background abundance (A_b) and intensity ratios (M_b), where

$$\begin{aligned} A_b &= \frac{1}{2}\log_2(R_bG_b) \text{ and} \\ M_b &= \log_2\left(\frac{R_b}{G_b}\right). \end{aligned}$$

Hereafter, foreground and background ratios refer to $\log_2$ ratios of intensities unless otherwise explicitly stated.

2.3 Simulation model

We assume an additive model stating that the true biological signal is the difference in the spot intensity and the latent background intensity. Let $\mathbf{\Theta}$ denote a 4-dimensional vector of parameters that represent the true state of nature for a single spot: $\mathbf{\Theta} = (\theta_A, \theta_{A_b}, \theta_M, \theta_{M_b})$, where θ_X is the parameter for the statistic X. Note that the true ratio of differential expression (θ_{M_s}) is known through $\mathbf{\Theta}$, that is

$$\theta_{M_s} = \log_2\left(\frac{\theta_R - \theta_{R_b}}{\theta_G - \theta_{G_b}}\right) = \log_2\left(\frac{2^{\theta_A + \frac{\theta_M}{2}} - 2^{\theta_{A_b} + \frac{\theta_{M_b}}{2}}}{2^{\theta_A - \frac{\theta_M}{2}} - 2^{\theta_{A_b} - \frac{\theta_{M_b}}{2}}}\right). \tag{2.1}$$

The assumption that background and biological signal add to equal the spot intensity needs further empirical verification.

For a given $\mathbf{\Theta}$, we simulate the expression of thousands of genes that vary around this truth without relying on distributional assumptions that are difficult to verify. SVS hybridizations are a natural choice since we are able to observe variability in differential expression across a range of abundance when the true differential expression is known to be absent. The algorithm for the simulation is outlined in Figure 1 and is repeated for each specification of $\mathbf{\Theta}$. Specifications of $\mathbf{\Theta}$ were chosen to cover a range of plausible values (see Section 3). Here, we describe the simulation for a fixed $\mathbf{\Theta}$.

Quantities of interest for a gene are simulated by independent random draws of observed normalized M and M_b from a SVS experiment. Sampling with replacement of M and M_b was restricted to deciles of spot abundance determined by θ_A as shown in Figure 1(a). The pair (M_i, M_{b_i}) denotes the i^{th} observation from two independent draws and need not correspond to an actual spot in the SVS hybridization. That is, M_i and M_{b_i} may correspond to the original foreground and background ratios of genes j and k, $j \neq k$. We

append the subscript $_{\theta_A}$ to M and M_b in Equations 2.2 and 2.3 to make their dependence on abundance explicit.

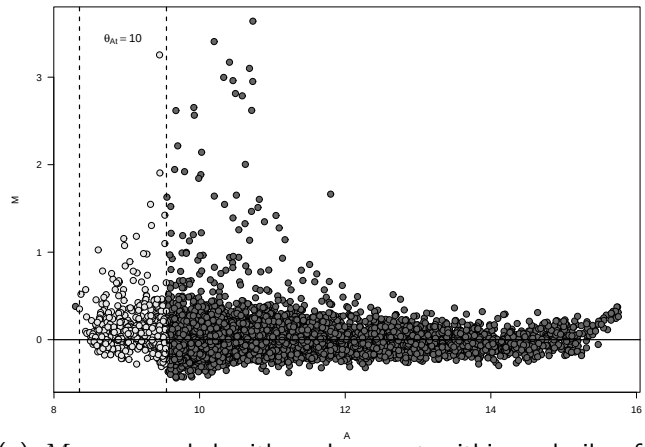
To assess the critical role of the relationship of M_{θ_A} to $M_{b\theta_A}$, we obtained the best fit line by linear regression:

$$M_{i\theta_A} = \beta_0 + \beta_1 M_{b_i\theta_A} + \epsilon_i. \tag{2.2}$$

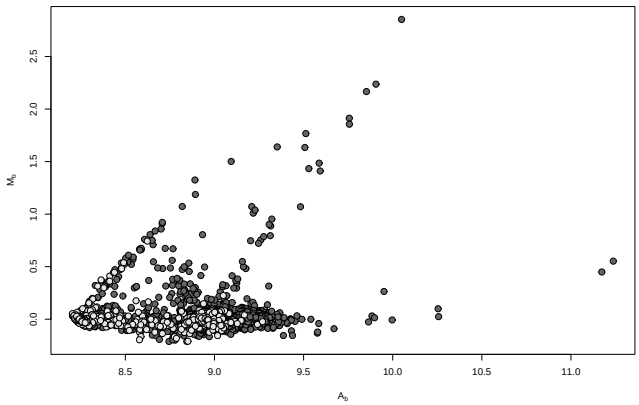
The simulation uses observed residuals directly so that it is not necessary to specify a distribution for these values. To manipulate the dependence of the foreground ratio to the background ratio, we parameterize the correlation of M_{θ_A} and $M_{b\theta_A}$ by ρ and vary this correlation by scaling the observed residuals in Equation 2.2 by a constant k such that

$$M_{i\theta_A} = \beta_0 + \beta_1 M_{b_i\theta_A} + k\epsilon_i, \quad \text{as shown in Figure 1(d).} \tag{2.3}$$

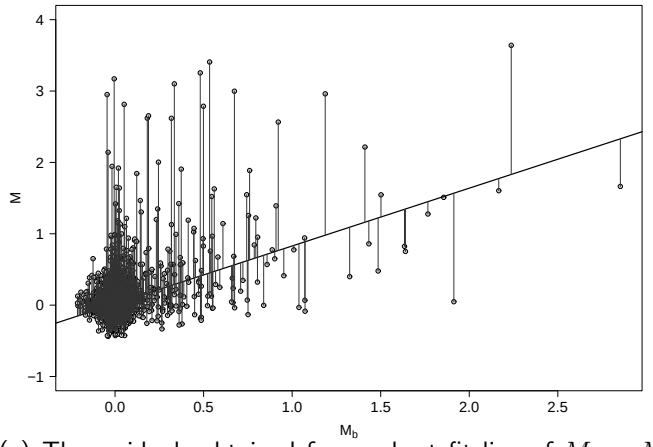
.



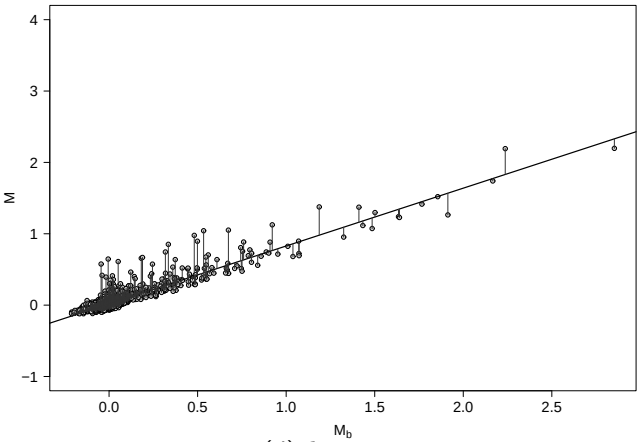
(a) M are sampled with replacement within a decile of A determined by θ_A.



(b) Highlighted in light gray are the spots in the first decile of A shown in Figure 1(a). We sample with replacement the M_b (vertical axis) of the gray spots.



(c) The residuals obtained from a best fit line of M on M_b are scaled by a constant k. Here, $k = 1$.



(d) $k = 0.2$

Figure 1: Top: MA plots following loess normalization of foreground (a) and background (b) for MCF7. Loess normalization of M_b in (b) was performed independently of the normalization of M. We simulate quantities of interest for a gene by independent random draws of M and M_b from their respective scatterplots. In this way, the distribution of the foreground and background ratios are dependent on the decile of abundance determined by θ_A. Bottom: The residuals from a linear regression of M on M_b are scaled to simulate different correlation structures of spot and background intensity ratios.

The average M_{θ_A} and $M_{b\theta_A}$ is zero. We simulate nonzero foreground ratios ($M^\dagger$) and background ratios ($M_b^\dagger$) by the following relationships:

$$\begin{aligned} M_i^\dagger &= \theta_M + M_{i\theta_A} \text{ and} \\ M_{b_i}^\dagger &= \theta_{M_b} + M_{b_i\theta_A}. \end{aligned} \tag{2.4}$$

Hence, the simulated foreground ratio, $M^\dagger$, is obtained from the adjusted residuals in Equation 2.3 by shifting the regression line by an amount given by the true foreground ratio, θ_M. To see this, we can rewrite Equation 2.4 as

$$\begin{aligned} M_i^\dagger &= \theta_M + \beta_0 + \beta_1 M_{b_i\theta_A} + k_\rho \epsilon_i \\ &= \theta_M + \beta_0 + \beta_1 \left(M_{b_i}^\dagger - \theta_{M_b} \right) + k_\rho \epsilon_i \\ &= \beta_0^* + \beta_1 \left(M_{b_i}^\dagger - \theta_{M_b} \right) + \epsilon_i^*. \end{aligned}$$

To summarize, we have simulated ratios of foreground that have an abundance-dependent distribution determined by the SVS experiment and whose dependence on background is parameterized by ρ. Calculations to obtain the simulated foreground and background intensities are straightforward:

$$\begin{aligned} \log_2(G_i^\dagger) &= \theta_A - \frac{M_i^\dagger}{2} \\ \log_2(R_i^\dagger) &= \theta_A + \frac{M_i^\dagger}{2} \\ \log_2(G_{b_i}^\dagger) &= \theta_{A_b} - \frac{M_{b_i}^\dagger}{2} \qquad (2.5) \\ \log_2(R_{b_i}^\dagger) &= \theta_{A_b} + \frac{M_{b_i}^\dagger}{2}. \qquad (2.6) \end{aligned}$$

The bias and variance with BS and NBS are given by the following relationships:

$$\begin{aligned}
\text{bias}_{BS} &= E\left[\log_2\left(\frac{R^\dagger - R_b^\dagger}{G^\dagger - G_b^\dagger}\right)\right] - \theta_{M_s} \\
&= E\left(M_s^\dagger\right) - \theta_{M_s} \\
\text{bias}_{NBS} &= E\left[\log_2\left(\frac{R^\dagger}{G^\dagger}\right)\right] - \theta_{M_s} \\
&= E\left(M^\dagger\right) - \theta_{M_s} \\
\text{variance}_{BS} &= \text{Var}\left[\log_2\left(\frac{R^\dagger - R_b^\dagger}{G^\dagger - G_b^\dagger}\right)\right] = \text{Var}\left(M_s^\dagger\right) \\
\text{variance}_{NBS} &= \text{Var}\left[\log_2\left(\frac{R^\dagger}{G^\dagger}\right)\right] = \text{Var}\left(M^\dagger\right).
\end{aligned}$$

Estimates of the bias, variance, and mean squared error (MSE) were obtained by averaging over 1000 simulations.

3 Results

We suggest two simple diagnostic plots to explore whether BS is needed: spatial images of background (logarithm scale) and scatterplots of M versus M_b. Figure 2 shows images of background from the Stratagene (row 1) and MCF7 (row 2) experiments. For Stratagene, $\log_2 R_b$ (plot 1) and $\log_2 G_b$ (plot 2) are comparable across most locations of the array and background is reasonably homogeneous with the notable exception of the lower right sector. The Spearman correlation coefficient for the M versus M_b scatterplot was relatively high (0.54). The high correlation of the foreground to background ratios suggests that a reduction in bias by BS may be achieved. By contrast, the spatial images of MCF7 background are more heterogeneous across channels and the Spearman correlation coefficient for the M versus M_b scatterplot is much lower (0.14). While background is reasonably homogeneous across the red and green channels for most regions of the array, a region in cell 2,1 and several regions in column 2 show more heterogeneity across channels. Whether the correlation of the ratios is also small within these subregions of the array is an important question that may influence both the bias and variability in the estimates of differential expression for a large number of genes in this region of the array. We further discuss the important issue of spatially-dependent correlations of the foreground and background ratios in Section 4. Nevertheless, it is not clear whether the reduction in bias achieved by BS in either experiment will offset the added variability. To see why, Figure 3 contains MA-plots that illustrate the variability in the foreground ratios following normalization with BS and NBS. The variability of the background ratios, M_s, is largest for low abundance genes. Clearly, NBS reduces the variability of the intensity ratios for low abundance genes (Figures 3(b) and 3(d)). More formally, we treat the decision of BS versus NBS as a trade off between bias and variance that can be addressed through simulation.

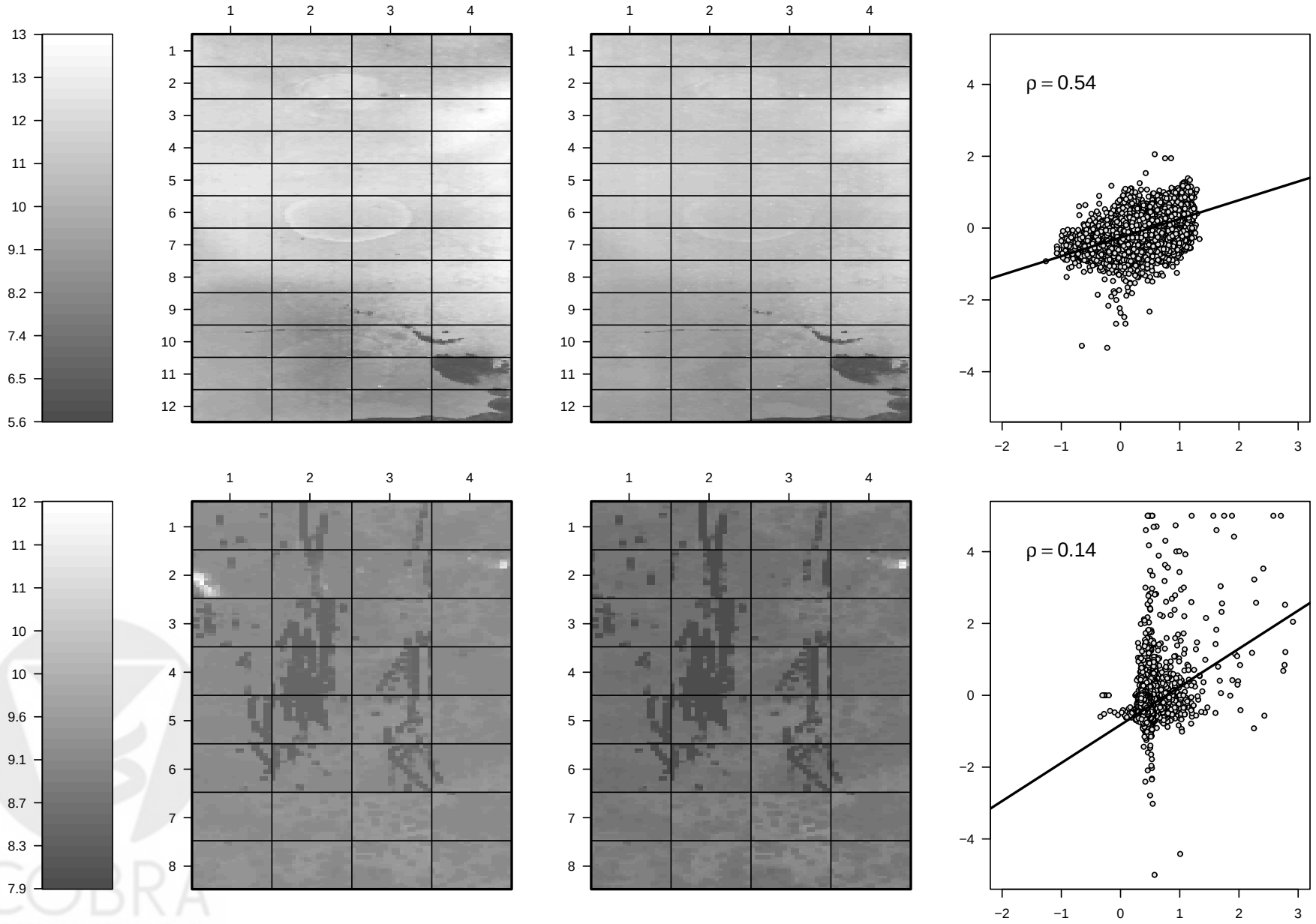
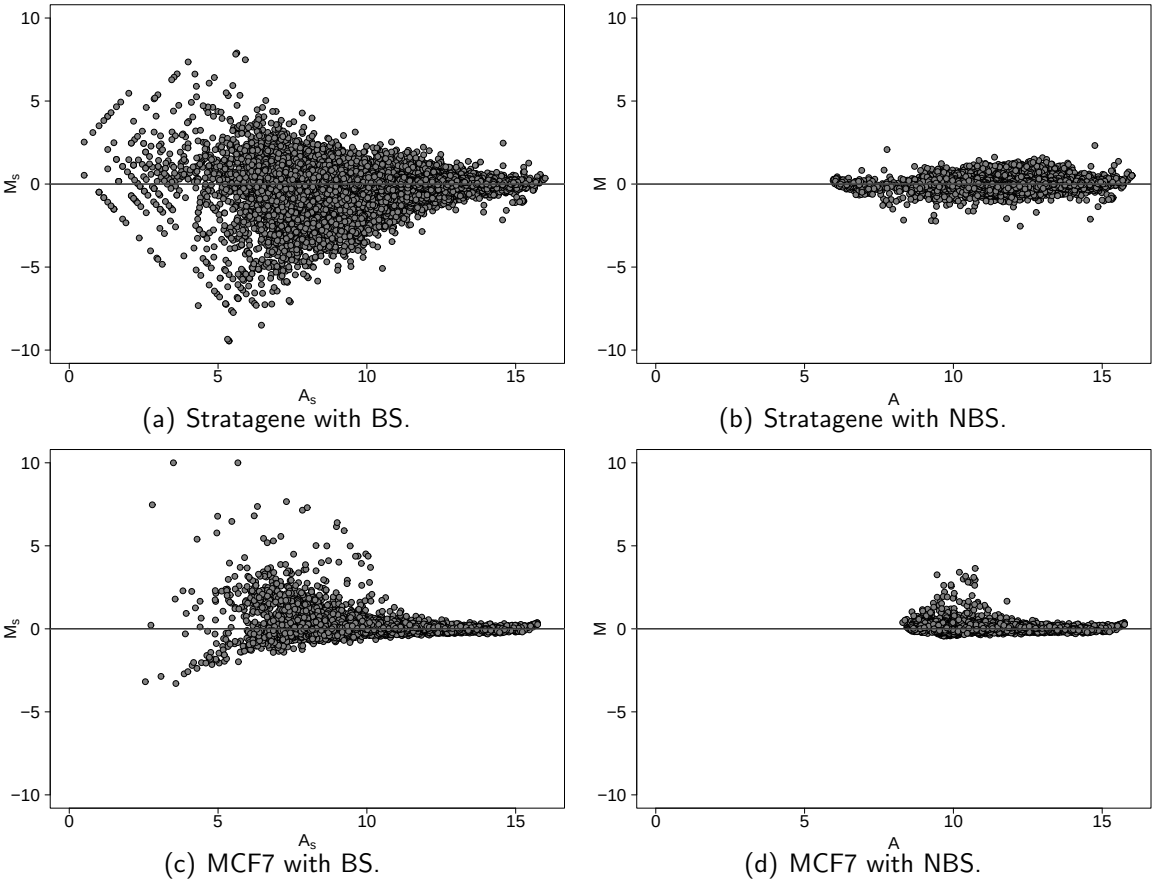


Figure 2: Diagnostic plots for exploring whether BS is needed for the Stratagene (row 1) and MCF7 (row 2) datasets. Spatial images of the pre-normalized $\log_2(R_b)$ and $\log_2(G_b)$ are plotted in columns 1 and 2, respectively. The third column shows M (vertical axis) versus M_b scatterplots for the respective experiments with the regression line and Spearman correlation (ρ) overplotted. M are truncated at 5 in the MCF7 scatterplot. Higher correlations of M and M_b in the Stratagene dataset suggest variation of the foreground ratios that is not driven by complementarity of probe and target sequences. We note substantial differences in the background intensities in cell 2,1 of the MCF7 experiment.

For the Stratagene dataset, For the MCF7 dataset,



(a) Stratagene with BS. (b) Stratagene with NBS.

(c) MCF7 with BS. (d) MCF7 with NBS.

Figure 3: MA scatterplots from two SVS microarrays following loess normalization with BS (left column) and NBS (right column). In truth, there is no differential expression. We observe substantial variability of the foreground ratios with BS. NBS is one method for reducing the variability of the foreground ratios of low abundance genes.

To determine when it is preferable to perform BS, we performed the simulation described in Section 2.3. The range of values specified for the parameters in the simulation is an important consideration and were chosen to illustrate the trade off in bias and variance, as well as to reflect empirically determined ranges in the SVS experiments. In particular, θ_A was chosen so that foreground ratios were sampled within deciles 1, 3, 5, and 7 of the observed A. Values for θ_{A_b} were determined by the median of the first and third quartiles of A_b.

Figures 4 and 5 show the bias-variance trade off for the simulations using the Stratagene and MCF7 data, respectively. Each panel plots the difference in estimates of the bias, variance, and MSE (vertical axis) using BS and NBS across a range of θ_{M_s} (horizontal axis) for a fixed θ_A (row) and fixed θ_{M_b} (column). Each of the three 16-panel plots in Figures 4 and 5 differ with respect to the correlation of $M^\dagger$ to $M_b^\dagger$ given by ρ. Shown here are correlations of 0.3, 0.2, and 0.1. Estimates of MSE using higher correlations (0.4) uniformly favored BS, whereas simulations using correlations lower than 0.1 (0.05) uniformly favored NBS (data not shown). The results shown here are for θ_{A_b} equal to the median of the first quartile of A_b. However, our findings were qualitatively similar using a value of θ_{A_b} in the third quartile of A_b (data not shown).

The first feature one might notice in the two plots with $\rho = 0.3$ and 0.2 is the concavity of the solid black line representing the difference in MSE. The concavity can be explained by observing that when the $\log_2$ ratio of the true signal is negative, we are likely to have a large bias in the estimate of differential expression if we do not subtract background ratios that are positive. The penalty in bias will be proportional to the correlation of the ratios of spot and background intensities, with higher correlations reflected by more negative lines for differences in bias and MSE. If the correlation is small (as in the plot with $\rho = 0.1$), the bias from NBS does not outweigh the cost of the added variability as

reflected by a positive MSE line.

Comparing Figure 4 to Figure 5, we observe similar trends but BS is generally more favorable in the simulation using the MCF7 data for $\rho = 0.2$ and higher, whereas in the Stratagene experiment NBS was preferable for $\rho = 0.2$. However, if we filtered the Stratagene data on flags generated by the imaging software and pixel-level correlations of red and green intensities, BS was also preferable for $\rho = 0.2$ and higher (data not shown). While there appears to be less variability in the foreground ratios after more stringent filtering, this further reduces the number of genes considered in the analysis by 40%. Finally, note that in each of the plots the BS decision is less critical as the ratio of θ_A/θ_{A_b} increases. For instance, in Figure 4 the largest difference in MSE for row 1 is roughly 20 fold greater than the largest difference in MSE for row 4.

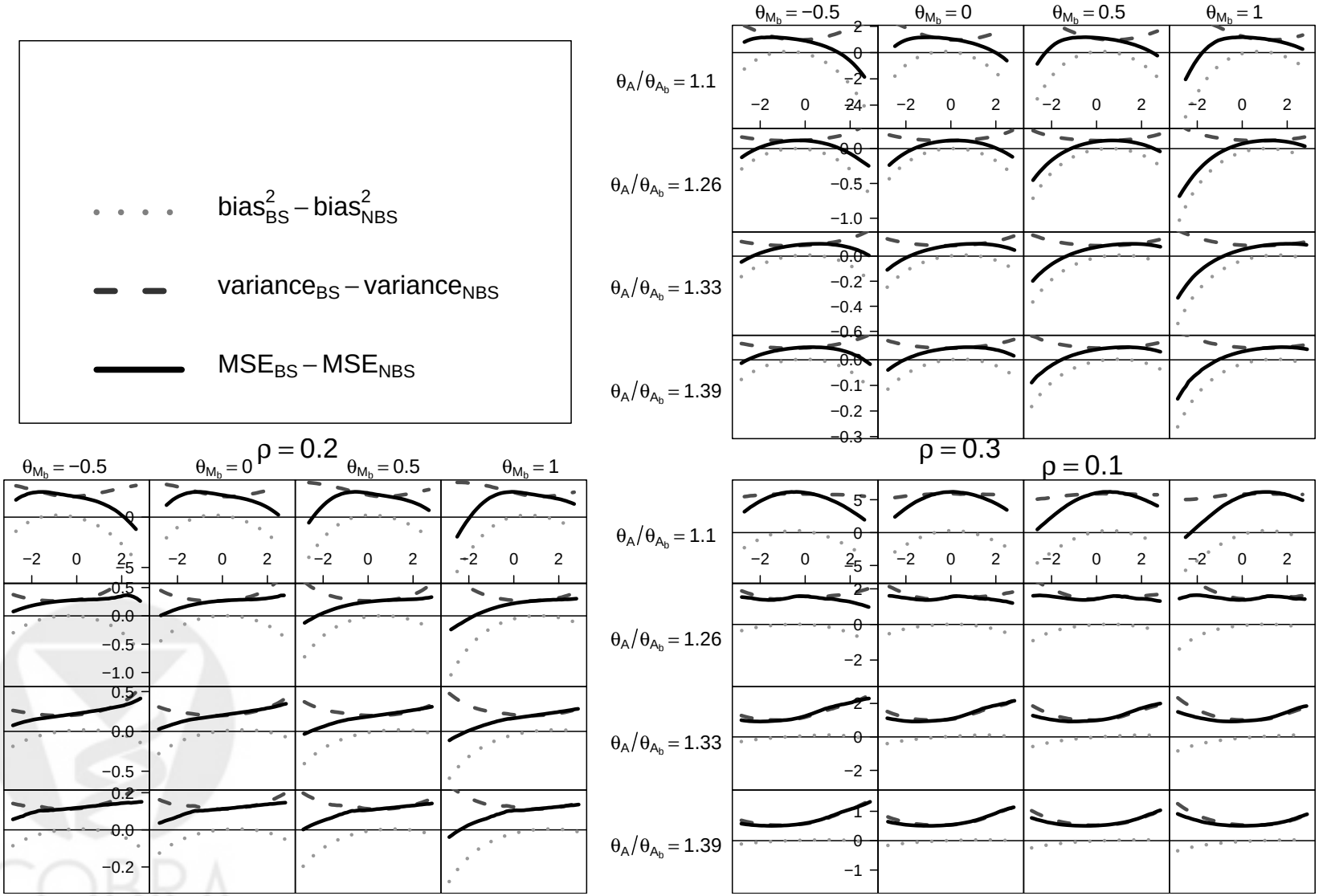


Figure 4: Simulation using the Stratagene SVS experiment. Each panel plots the difference in bias, variance, and MSE (vertical axis, fixed by row) for BS versus NBS across a range of θ_{M_s} (horizontal axis) for a fixed θ_A (row) and fixed θ_{M_b} (column). The 3 plots differ by ρ. For each of the 3 plots, row 1 shows the trade-off between bias and variance when the ratio of θ_A/θ_{A_b} is small, with increasing ratios of θ_A/θ_{A_b} down a column and increasing θ_{M_b} across columns. BS is the preferred method when the correlation of $M^\dagger$ to $M_b^\dagger$ is relatively high since the change in MSE is below zero for most panels, whereas NBS is preferable for smaller correlations. The bias-variance trade-off is most evenly balanced when the correlation is 0.30.

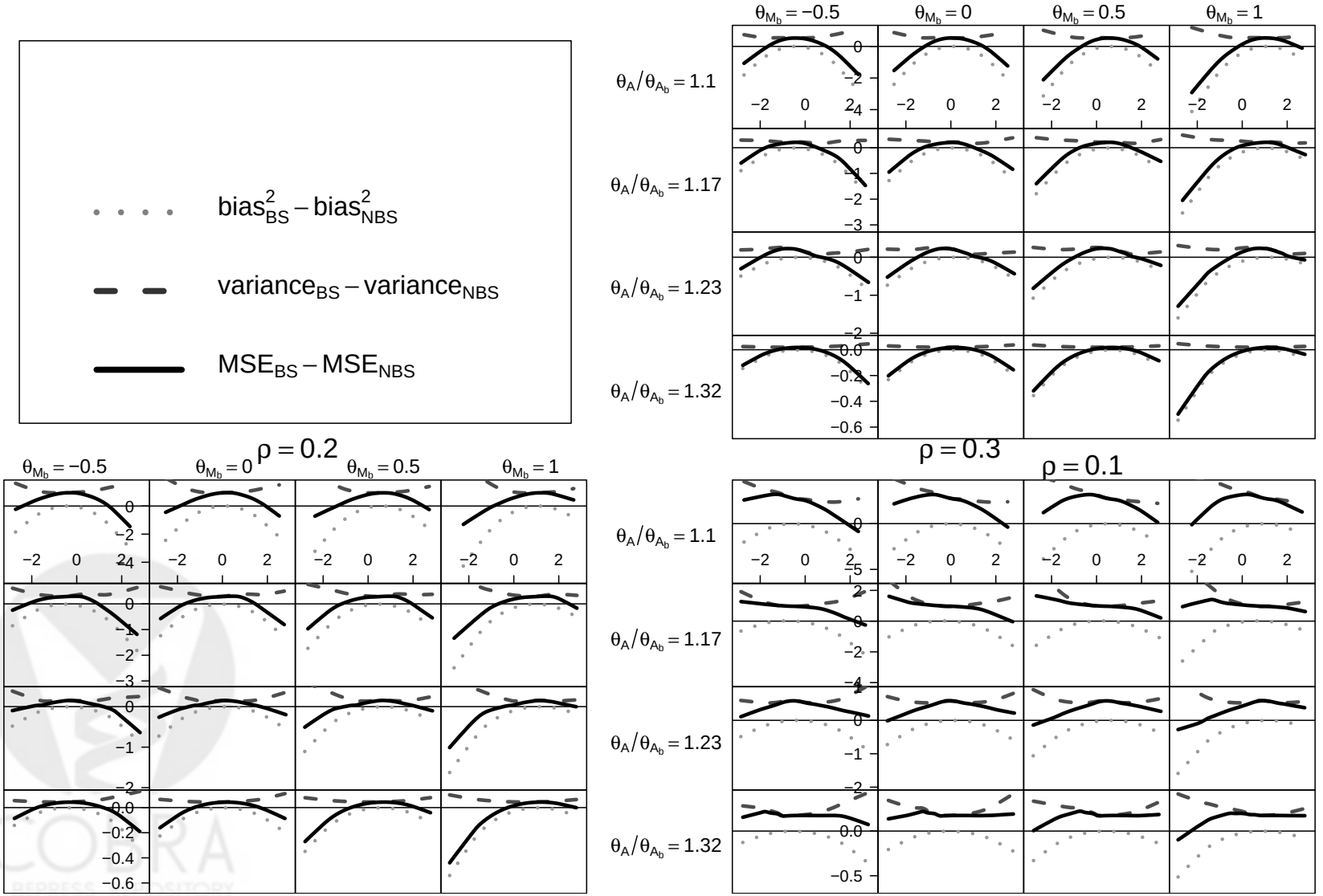


Figure 5: Simulation using the MCF7 cell line SVS. Each panel plots the difference in bias, variance, and MSE (vertical axis, fixed by row) for BS versus NBS across a range of θ_{M_s} (horizontal axis) for a fixed θ_A (row) and fixed θ_{M_b} (column). In general, BS is preferable when the correlation of foreground to background ratios is greater than 0.2.

4 Discussion

Images of background fluorescence together with scatterplots of ratios of foreground and background are useful diagnostics for aiding the decision of whether to subtract estimates of local background. Through simulation, we show the bias-variance trade off of BS over a range of experimental conditions observed in practice.

Because the variability of gene expression in cDNA microarrays is known to be dependent on spot abundance [17, 1], our simulation captures abundance-dependent variability from two SVS hybridizations where the true differential expression is known to be absent. In this way, we avoid specifying a parametric model and simulate genes having a range of possible true differential expressions with varying levels of abundance and background intensity ratios. Additionally, we simulated foreground ratios with varying degrees of correlation to background ratios. Figures 4 and 5 show that BS is less favorable in terms of MSE across a range of possible truths for differential expression when the correlation of background and spot intensity ratios was low (0.1 and less). Conversely, high correlation (0.3 and greater) of foreground to background ratios favors BS and may indicate that background is not spot-localized, or that appreciable non-sequence based fluorescence occurs within regions of hybridization. Pre-processing procedures that do not BS are penalized by the large bias in these instances.

As the correlation of foreground to background ratios is likely to vary across locations of the microarray, a useful diagnostic is to calculate the correlation of foreground to background ratios for each cell of the microarray grid. For instance, in Figure 2 we observed substantial heterogeneity across the images for red and green background in cell 2,1. As noted previously, if the foreground and background ratios are highly correlated failure to BS can result in a large bias. The M versus M_b scatterplot for the MCF7

experiment in Figure 2 has an overall low correlation (0.14) and the recommendation from the simulation is NBS. Although correlations of foreground and background ratios were very low for most cells of the array (Figure 6), the ratios of foreground and background were well correlated in cell 2,1 and for several cells in column 2. Because the distribution of the correlations in Figure 6 is predominantly less than 0.2, a combined approach that performs BS for the 7 cells (of 32) that are well-correlated, and peforms NBS for the remaining cells may be reasonable. However, if the distribution of the correlations is less polarized to one methodology, such a combined approach would not be suitable. Methods that perform BS as a smooth function of ρ is a future direction of this work and could be evaluated using the simulation described here.

Microarrays have different sources of technological variation that may arise at any step during the experiment, including RNA preparation, printing of the microarray, and imaging of the hybridized samples. Our simulation is based on two SVS hybridizations and as such may have noise that differs from other microarrays. That the results were qualitatively similar across two SVS experiments produced by different labs, different biological samples, and different imaging software suggests that our simulation is not overly sensitive to the specific technological variability in these two experiments. Additionally, while we performed our analysis with cDNA microarrays, this work is relevant and can be easily adapted to other other 2-channel microarray platforms, such as Agilent [13].

The findings presented here are consistent with others who have used different methods to evaluate BS. For instance, Qin *et al.* compare BS to NBS in four microarray experiments with spike-in genes (genes inserted at known ratios) [20]. BS was inferior to NBS in each of the four microarray experiments, largely because of the variability in the $\log$ intensity ratios of the low-mid abundance genes.

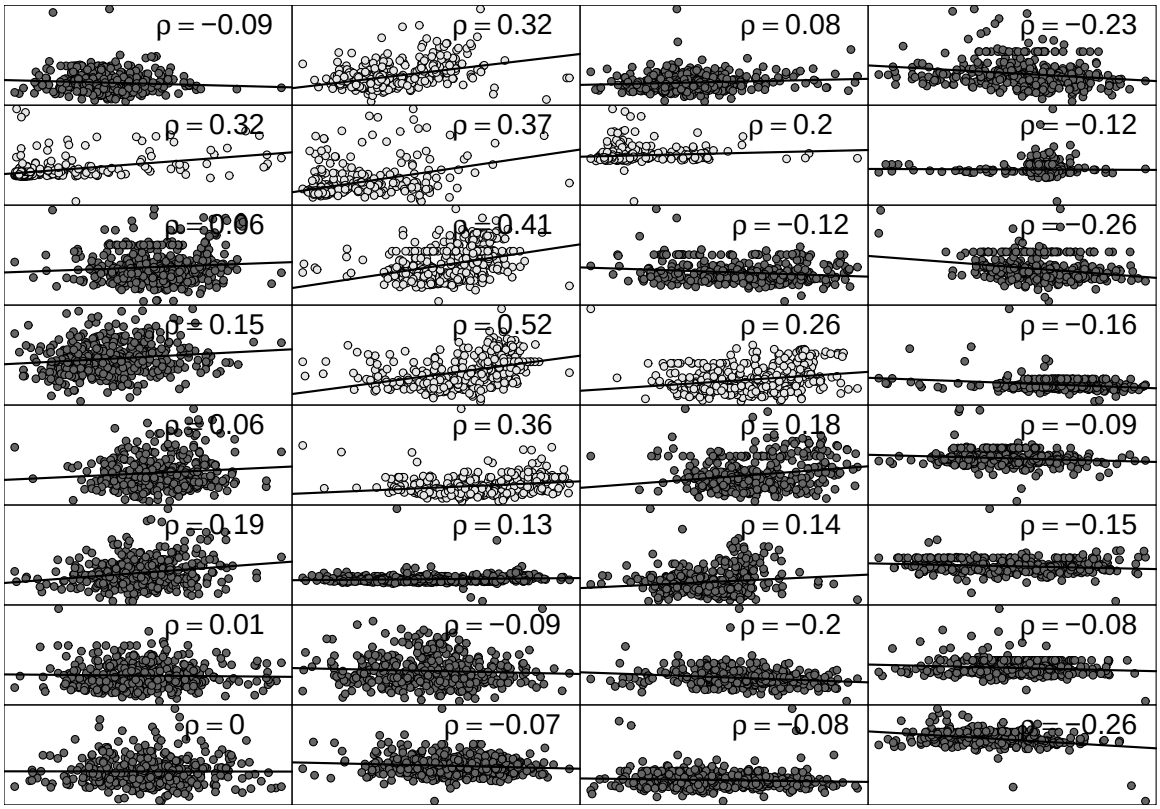


Figure 6: The Spearman correlation of the pre-normalized M (vertical axis) versus M_b scatterplots was calculated for each cell of the microarray for the MCF7 dataset. The cells with spheres highlighted in light gray indicate that the correlation of foreground to background ratios is 0.2 or higher and are regions where BS is recommended by the simulation (Figure 5). Note that in cell 2,1 and for several cells in column 2 the recommendation is to BS (see also cell 2,1 and column 2 of Figure 2). The regression lines are overplotted.

Our simulation shows the relationship of bias, variance, and MSE for intensity-dependent normalization procedures performed with and without BS across a range of simulated differential expressions. The correlation of foreground to background ratios is an important consideration before subtracting background fluorescence. M versus M_b scatterplots are a useful diagnostic to locate regions of the array where BS may be inappropriate. A normalization methodology tailored to a particular microarray experiment will put researchers in a better position to identify differentially expressed genes.

Acknowledgments

We thank John Berger for making the data for the MCF7 cell line publicly available. We thank Leslie Cope for his comments and suggestions regarding this manuscript. Work supported by NSF grant NCIP30CA06973.

References

[1] K.A. Baggerly, K.R. Coombes, K.R. Hess, D.N. Stivers, L.V. Abruzzo, and W. Zhang. Identifying differentially expressed genes in cDNA microarray experiments. *Journal of Computational Biology*, 8:639–659, 2001.

[2] John A Berger, Sampsa Hautaniemi, Anna-Kaarina Jrvinen, Henrik Edgren, Sanjit K Mitra, and Jaakko Astola. Optimized LOWESS normalization parameter selection for DNA microarray data. *BMC Bioinformatics*, 5(1):194, Dec 2004.

[3] C. S. Brown, P. C. Goodwin, and P. K. Sorger. Image metrics in the statistical analysis of DNA microarray data. *Proc Natl Acad Sci U S A*, 98(16):8944–9, Jul 2001.

[4] Y Chen, E Dougherty, and M Bittner. Ratio-based decisions and the quantitative analysis of cDNA micro-array images. *Journal of Biomedical Optics*, 2:364–374, 1997.

[5] Y. Chen, V. Kamat, E.R. Dougherty, M.L. Bittern, P.S. Meltzer, and J.M. Trent. Ratio statistics of gene expression levels and applications to microarray data analysis. *Bioinformatics*, 18:1207–1215, 2002.

[6] W.S. Cleveland. Robust locally weighted regression and smoothing scatterplots. *Journal of American Statistics Association*, 74:829–836, 1979.

[7] W.S. Cleveland. Lowess: A program for smoothing scatterplots by robust locally weighted regression. *The American Statistician*, 35:54, 1981.

[8] S. Dudoit and J.Y.H. Yang. Bioconductor R packages for exploratory analysis and normalization of cDNA microarray data. In G. Parmigiani, E.S. Garrett, R.A. Irizarry, and S.L. Zeger, editors, *The Analysis of Gene Expression Data: Methods and Software*, New York, 2003. Springer Verlag.

[9] M. R. Fielden, R. G. Halgren, E. Dere, and T. R. Zacharewski. GP3: GenePix post-processing program for automated analysis of raw microarray data. *Bioinformatics*, 18(5):771–3, May 2002.

[10] Robert Gentleman. BioConductor: open source software for bioinformatics. `http://www.bioconductor.org`, 2003.

[11] J Gollub, CA Ball, G Binkley, J Demeter, DB Finkelstein, JM Hebert, T Hernandez-Boussard, H Jin, M Kaloper, JC Matese, M Schroeder, PO Brown, D Botstein, and G Sherlock. Nucleic acids res. *The Stanford Microarray Database: Data access and quality assessment tools*, 31(1):94–96, 2003.

[12] G Hardiman. Microarray technologies—an overview. *Pharmacogenomics*, 3(3):293–7, 2002.

[13] T.R. Hughes, M. Mao, A.R. Jones, J. Burchard, M.J. Marton, K.W. Shannon, S.M. Lefkowitz, M Ziman, J.M. Schelter, M.R. Meyer, and et al. Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nature Biotechnology*, 19(4):342–347, 2001.

[14] Ross Ihaka and Robert Gentleman. R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5(3):299–314, 1996.

[15] Charles Kooperberg, Thomas G Fazzio, Jeffrey J Delrow, and Toshio Tsukiyama. Improved background correction for spotted DNA microarrays. *Journal of Computational Biology*, 2002.

[16] M. Juanita Martinez, Anthony D. Aragon, Angelina L. Rodriguez, Jose M. Weber, Jerilyn A. Timlin, Michael B. Sinclair, David M. Haaland, and Margaret Werner-Washburne. Identification and removal of contaminating fluorescence from commercial and in-house printed DNA microarrays. *Nucl. Acids. Res.*, 31(4):e18–, 2003.

[17] M. A. Newton, C. M. Kendziorski, C. S. Richmond, F. R. Blattner, and K. W. Tsui. On differential variability of expression ratios: Improving statistical inference about gene expression changes from microarray data. *Journal of Computational Biology*, 8:37–52, 2001.

[18] T Park, SG Yi, SH Kank, SY Lee, YS Lee, and R Simon. Evaluation of normalization methods for microarray data. *BMC Bioinformatics*, 2003.

[19] Giovanni Parmigiani, Elizabeth S Garrett, Rafael A Irizarry, and Scott L Zeger. *The Analysis of Gene Expression Data: Methods and Software*. Springer, New York, 2003.

[20] Li-Xuan Qin, Kathleen F. Kerr, and Contributing Members of the Toxicogenomics Consortium. Empirical evaluation of data transformations and ranking statistics for microarray analysis. *Nucleic Acids Research*, 32(18):5471–5479, 2004.

[21] John Quackenbush. Microarray data normalization and transformation. *Nature Genetics*, 32S:496–501, December 2002.

[22] Mark Schena. *Microarray Biochip Technology*. BioTechniques Press, Westborough, MA, 2000.

[23] Johannes Schuchhardt, Dieter Beule, Arif Malik, Eryc Wolski, Holger Eickhoff, Hans Lehrach, and Hanspeter Herzel. Normalization strategies for cDNA microarrays. *Nucl. Acids. Res.*, 28(10):e47–, 2000.

[24] R. M. Simon, E.L. Korn, L.M. McShane, M.D. Radmacher, G. W. Wright, and Y. Zhao. *Design and Analyis of DNA Microarray Investigations*. Springer, New York, 2003.

[25] E M Southern. DNA microarrays. History and overview. *Methods in Molecular Biology*, 170:1–15, 2001.

[26] T. P. Speed, editor. *Statistical Analysis of Gene Expression Microarray Data*. Chapman and Hall, London, 2003.

[27] Ivana V Yang, Emily Chen, Jeremy P Hasseman, Wei Liang, Bryan C Frank, Shuibang Wang, Vasily Sharov, Alexander I Saeed, Joseph White, Jerry Li, Norman H Lee, Timothy J Yeatman, and John Quackenbush. Within the fold: assessing differential expression measures and reproducibility in microarray assays. *Genome Biol*, 3(11):research0062, Oct 2002.